Published in final edited form as:
Dev Biol. 2007 June 15; 306(2): 860–869.

Cis-regulatory control of the *nodal* gene, initiator of the sea urchin oral ectoderm gene network

Jongmin Nam[a], **Yi-Hsien Su**[a], **Pei Yun Lee**[a], **Anthony J. Robertson**[b], **James A. Coffman**[b], and **Eric H. Davidson**[a,*]

a*Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA*

b*Mount Desert Island Biological Laboratory, Salisbury Cove, ME 04672, USA*

Abstract

Expression of the *nodal* gene initiates the gene regulatory network which establishes the transcriptional specification of the oral ectoderm in the sea urchin embryo. This gene encodes a TGFβ ligand, and in *Strongylocentrotus purpuratus* its transcription is activated in the presumptive oral ectoderm at about the 30-cell stage. Thereafter Nodal signaling occurs among all cells of the oral ectoderm territory, and nodal expression is required for expression of oral ectoderm regulatory genes. The *cis*-regulatory system of the *nodal* gene transduces anisotropically distributed cytoplasmic cues that distinguish the future oral and aboral domains of the early embryo. Here we establish the genomic basis for the initiation and maintenance of *nodal* gene expression in the oral ectoderm. Functional *cis*-regulatory control modules of the *nodal* gene were identified by interspecific sequence conservation. A 5′ *cis*-regulatory module functions both to initiate expression of the *nodal* gene and to maintain its expression by means of feedback input from the Nodal signal transduction system. These functions are mediated respectively by target sites for bZIP transcription factors, and by SMAD target sites. At least one SMAD site is also needed for the initiation of expression. An intron module also contains SMAD sites which respond to Nodal feedback, and in addition acts to repress vegetal expression. These observations explain the main features of *nodal* expression in the oral ectoderm: since the activity of bZIP factors is redox sensitive, and the initial polarization of oral vs aboral fate is manifested in a redox differential, the bZIP sites account for the activation of nodal on the oral side; and since the immediate early signal transduction response factors for Nodal are SMAD factors, the SMAD sites account for the feedback maintenance of nodal gene expression.

Keywords

Nodal; Oral ectoderm; Gene regulatory network; Community effect; TGF-beta; bZIP; SMAD; Sea urchin; Positive feedback regulation; *cis*-regulatory analysis

Introduction

Expression of the *nodal* gene was identified as the initial transcriptional event in oral ectoderm specification in *Paracentrotus lividus* by Duboc et al. (2004), and similarly in *Strongylocentrotus purpuratus* by Flowers et al. (2004). Not only is *nodal* expression the earliest known oral ectoderm-specific transcriptional activity, but observations on oral

*Corresponding author: Eric H. Davidson, Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA, Fax: +1 626 793 3047, E-mail address: davidson@caltech.edu (E.H. Davidson)

ectoderm marker genes demonstrate that if this expression is blocked, oral ectoderm differentiation fails to occur (Duboc et al., 2004). In *S. purpuratus* the *nodal* gene is activated by 5th cleavage, which is the stage at which oral vs aboral founder cell lineages have all separated, according to lineage tracing studies (Cameron et al., 1990). In current studies to be presented elsewhere, we determined the architecture of the gene regulatory network underlying specification of the embryonic oral ectoderm territory. The activation of all the initial tier of regulatory genes in this network, and hence indirectly of all the more downstream genes as well, depends on reception of the Nodal signal by the cells of the presumptive oral ectoderm: expression of these genes fails if nodal expression is blocked by treatment with morpholino substituted antisense oligonucleotides (MASO). Thus the initial key to the genomic regulatory code for oral ectoderm specification must lie in the *cis*-regulatory control system which causes the activation of the *nodal* gene, and maintains its expression in the oral ectoderm.

Several clues from earlier work indicate possible inputs into the *nodal cis*-regulatory system. First, Coffman and Davidson (2001) and Coffman et al. (2004) showed that oral-aboral polarity is initially reflected by a redox gradient in early embryos, which is induced by polarized distribution of maternal mitochondria. Furthermore, experimentally perturbed redox polarization causally results in oral ectoderm specification. Further downstream, the redox sensitive kinase P38 is likely to be involved in transduction of the initial anisotropy into a polarized transcriptional regulatory state (Bradham and McClay, 2006). This suggests that the nodal *cis*-regulatory system might respond to a redox sensitive transcription factor. Second, since all cells of the oral ectoderm express *nodal* and all cells of this territory also receive the Nodal signal, it is a reasonable supposition that the *nodal* gene might be controlled by a feedback mediated through the immediate early response factor that is activated by reception of the Nodal signal. Expression of *nodal* is regulated in this way in vertebrates, though in very different developmental contexts (Schier, 2003). Third, evidence from microsurgical experiments implies that some as yet unknown vegetal signal may be required for specification of the overlying oral ectoderm, since isolated animal halves fail to develop oral ectoderm, as do embryos in which endomesodermal specification is blocked by interference with β-catenin nuclearization (Wikramanayake et al., 1995;Wikramanayake et al., 1998;Wikramanayake and Klein, 1997). Fourth, the pattern of *nodal* expression could imply repression in apical and vegetal domains, as the gene is silent in these polar regions even on the oral side of the embryo.

In the following we describe the isolation and functional characterization of the relevant *nodal cis*-regulatory modules. Mutational analysis of transcription factor target sites within these modules reveals the *cis*-regulatory logic by which this initial step in oral ectoderm specification is controlled.

Materials and methods

Quantitative polymerase chain reaction (QPCR)

Embryos grown at 14°C were sampled at multiple time points (hours post fertilization, hpf) and total RNAs from these samples were extracted using RNeasy mini kit (Qiagen, Valencia, CA). After DNase I treatment, < 1 μg of total RNA was used for cDNA synthesis using iScript cDNA synthesis kit following manufacturer's instruction (Bio-Rad laboratories, Hercules, CA). 2.5 μl of five-fold diluted cDNA pool was further used for each QPCR using iTaq SYBR Green Supermix with ROX kit (Bio-Rad laboratories, Hercules, CA). Numbers of transcripts were estimated following Wong and Medrano (2005), with some modifications (Materna et al., 2006). Ubiquitin was used as an internal standard for estimating the number of transcripts per embryo (Materna et al., 2006).

For experiments on microinjected embryos, about 150 embryos were sampled for each time point. AllPrep DNA/RNA mini kit (Qiagen, Valencia, CA) was used to extract genomic DNA

and total RNA simultaneously. The number of GFP or RFP transcripts was normalized to the number of DNA copies incorporated by using the method of Revilla-i-Domingo et al. (2004). The *nodal* sequence was used as standard for estimating the number of GFP or RFP DNA molecules incorporated.

The primers used for QPCR were *nodal* forward primer, *nodal* reverse primer, *ubiquitin* forward primer, *ubiquitin* reverse primer, GFP forward primer, GFP reverse primer, RFP forward primer, and RFP reverse primer. See supplementary table S1 for the sequences of primers.

Screening BAC clones and comparative sequence analysis

BAC clones containing the *nodal* locus from *S. purpuratus* (~180 kb) and *L. variegatus* (~130 kb) were obtained by screening arrayed libraries using ^{32}P-labeled probes generated from full length SpNodal cDNA (Coffman et al., 2004). A BAC clone containing *Lv-nodal* was sequenced and subsequently assembled.

The *Lv-nodal* BAC sequence obtained above and genomic sequence of *Sp-nodal* (~17 kb) obtained from the sea urchin genome project (Sea Urchin Genome Sequencing et al., 2006) were compared using FamilyRelation II (Brown et al., 2005). Several identity cutoffs from 70% to 90% for a 50 bp sliding window were applied, and conserved noncoding regions of ≥70% sequence identity were reported.

Generation of reporter gene constructs

Each of the conserved modules was fused to a GFP construct harboring the *endo16* basal promoter (Arnone et al., 1997) by fusion PCR (Yon and Fried, 1989). The primers used for amplifying each module from *Sp-nodal* BAC were 5P-module forward, 5P-module reverse, INT-module forward, INT-module reverse, 3P-module forward, and 3P-module reverse. See supplementary table S1 for the sequences of primers. The Expand high fidelity PCR system (Roche Diagnostics GmbH, Mannheim, Germany) was used for every fusion PCR reaction conducted in this study.

The *nodal* GFP-BAC knock-in construct was generated following the method of Lee et al. (2001) with minor modifications. The *Sp-nodal* BAC chosen (clone # F21-7) contains the entire 17 kb-long genomic region used for the FamilyRelation analysis. The first 159 bp of the nodal ORF was replaced by a GFP cDNA by homologous recombination. The primers used for amplifying GFP-Kanamycin cassette were GFP forward and Kanamycin reverse. We also generated an RFP knock-in construct by fusion PCR. The primers used for this fusion PCR were RFP forward and RFP reverse. See Supplementary Table S1 for the sequences of primers.

The primers used for amplifying different regions of the construct (see Fig. 2) were P1 forward, P2 forward, P3 forward, P3 reverse, P3-2 forward, P4 reverse, P5 forward, P5 reverse, P6 reverse, P7 reverse, P8 forward, P9 reverse, and P6-P8. See Supplementary Table S1 for the sequences of primers.

To mutate putative bZIP sites and putative SMAD sites, we changed four nucleotide core sequences within sites predicted by MatInspector software (Cartharius et al., 2005), using PCR primers containing the desired mutations. For each mutant construct, we first amplified from the nodal-GFP-BAC a fragment flanked by the two nearest target sites of our interest. We then fused corresponding fragments to generate necessary mutant constructs by PCR. If more than three mutations were needed in one construct (four or more fragments to be fused), we introduced new mutations by PCR using the mutant construct of other sites as template. The primers used for generating mutants were bZIP-1 forward, bZIP-1 reverse, bZIP-2 forward, bZIP-2 reverse, SMAD-1 forward, SMAD-1 reverse, SMAD-2 forward, SMAD-2 forward,

SMAD-3 forward, SMAD-3 reverse, SMAD-4 forward, and SMAD-4 reverse. Sites were numbered from 5′ to 3′ orientation in the 5P- and INT-modules. See Supplementary Table S1 for the sequences of primers.

Microinjection

Microinjection was conducted as described (Arnone et al., 2004). *nodal* MASO (GeneTools, Philomath, OR) was injected at 30 μM. A similar amount of a control MASO was also injected per egg. The sequence of the nodal MASO used was TGCATGGTTAAAAGTCCTTAAAAAT.

PCR products were purified using QIAquick PCR purification kit (Qiagen, Valencia, CA), and BAC DNAs were linearized with *Not*I digestion and were filter purified. For DNA injection, we estimate that about 500 molecules per embryo were injected for single construct, and about 250 molecules each per embryo for two different constructs.

Results

Temporal and spatial expression of the endogenous nodal gene

Transcripts of the *nodal* gene were measured by QPCR from fertilization to the 24 h blastula stage. As shown in Fig. 1, expression was not detectable until 6 h, about 5th cleavage. This is similar to the results of Duboc et al. (2004) for *Paracentrotus*, but we do not confirm the report of Flowers et al. (2004) that the *nodal* gene is represented by maternal mRNA in *S. purpuratus*. By 7 h there are ~100 molecules of *nodal* mRNA/embryo. The rate of transcription increases, and by 9 h the approximately steady state content of ~600 molecules/embryo is attained (Fig. 1). This level persists at least out to 48 h. Note that absolute level of expression of *nodal* mRNA in control embryos varies by about two fold between different batches of embryos.

We also carried out whole mount in situ hybridization (WMISH) to determine the spatial distribution of nodal transcript at 18 h and 24 h blastula stages. The results were the same as those of Duboc et al. (2004) and Flowers et al. (2004) in displaying nodal transcripts in about one third of the ectodermal area, i.e., the future oral ectoderm, excluding both the apical and vegetal poles (data not shown).

Positive feedback input and a model for temporal expression of nodal

The possibility raised in Introduction that *nodal* transcription is maintained by positive feedback regulation would mean that in each oral ectoderm cell, reception of the Nodal ligand and activation of the Nodal signal transduction system causes stimulation of *nodal* transcription. To test this we blocked Nodal translation by injection into eggs of a morpholino substituted antisense oligonucleotide (MASO), and then measured quantitative output of the endogenous *nodal* gene as in Fig. 1. Figure 2A shows that *nodal* MASO indeed sharply depresses *nodal* transcript levels. Moreover, significant down regulation is obvious at as early as 8 h, suggesting that the positive feedback begins to operate soon after the initial activation of the gene. The same result was consistently observed with very little variation (see figure legend). Figure 2A indicates that the initial activation of the gene accounts for no more than ~30 molecules of nodal mRNA/embryo, as estimated from the level in embryos bearing the nodal MASO, and this could be an overestimate if the MASO block is at all leaky.

A simple model for the temporal expression of *nodal* is shown in Fig. 2B, in which the total measured output is the sum of the initial input and expression due to feedback reinforcement. But since Nodal is a secreted ligand, even if it diffuses only to the immediately adjacent cell in these embryos (Yaguchi et al., 2007), this model raises the question why expression does

not spread around the embryo to include the aboral ectoderm. The answer to this question, to which we return below, could depend either on an off-the-DNA interference with Nodal signaling in the aboral ectoderm, or on characteristics of the *nodal cis*-regulatory system, or both.

cis-Regulatory modules and their spatial activities

In order to identify the *cis*-regulatory modules controlling *nodal* expression we compared *S. purpuratus* (~17 kb) and *Lytechinus variegatus* (~150 kb) genomic sequences containing the *nodal* locus, using the FamilyRelations II program (Brown et al., 2005). The underlying assumption of such analysis is that functional constraints on noncoding regions are reflected in decreased evolutionary rates of divergence. Figure 3A shows the locations of three relatively conserved noncoding regions of ≥70% sequence identity within a 50 bp sliding window. These are in the 5′-upstream proximal region (5P-module), in the intron (INT-module), and in the 3′-downstream region (3P-module). To determine the transcriptional activity of the three conserved regions, each was fused to a GFP construct harboring the *endo16* basal promoter (EpGFP hereafter, see Figure 3B and Arnone et al., 1997). The constructs were injected into fertilized eggs and GFP expression was scored at 24 h, a time by which the oral and aboral territories have been specified, and the auto-fluorescence of early stage embryos is much reduced. Results are summarized in Table 1. The 5P-EpGFP and INT-EpGFP constructs displayed GFP signals in 55% and 43% of observed embryos, respectively, while no GFP signal was produced by the 3P-EpGFP construct. Both the 5P-EpGFP and the INT-EpGFP constructs were active preferentially in ectoderm (Table 1).

We created a second generation of expression constructs which incorporated the endogenous basal promoter of the nodal gene in place of the *endo16* promoter of the EpGFP constructs. These constructs, displayed in Figs. 3C and D, were based on GFP BAC knock-ins (KI constructs), produced by in vitro recombination as described in Materials and methods. Red fluorescent protein (RFP) KI constructs were also built. The endogenous basal promoter is included in a DNA fragment that extends from -136 bp (P3-2 in Fig. 3A) to the transcription start site (Flowers et al., 2004); this fragment is required for the activity of the INT-module but by itself does not drive any expression of GFP, as measured by QPCR (data not shown). A 5P-module construct (5P-KI) was produced by amplifying the region of the GFP BAC from primers P2 to P4, and an INT construct (INT-KI) was similarly generated as the amplicon extending from P3-2 to P7 (Fig. 3C). Additional constructs made by the same strategy are shown in Figs. 3C and D.

The spatial activities of 5P-KI and INT-KI are compared to those of the whole Nodal-GFP BAC in the co-injection experiment of Fig. 4. The premises of this demonstration are that all injected DNAs ligate together in vivo and incorporate into the same cells of the embryo, as was demonstrated by Livant et al. (1991); and that the BAC, which extends from > 40 kb upstream to >15 kb downstream of the nodal exons, would include the complete regulatory system. That the GFP BAC indeed expresses correctly, i.e., comparably to the endogenous *nodal* gene, was shown by observations at later times when the oral ectoderm can be easily distinguished from aboral ectoderm morphologically (data not shown). The 24 h embryo displayed in Fig. 4A developed from an egg coinjected with the GFP BAC and a version of 5P-KI equipped with an RFP reporter: the expression of GFP and RFP are coincident, thus demonstrating that 5P-KI expresses just as does the whole regulatory system carried in the BAC. Figure 4B shows that INT-KI also expresses co-incidentally with 5P-KI. Therefore both the 5P- and INT-modules are expressed in the presumptive oral ectoderm.

Quantitative assessment of the spatial expression of 5P-KI and INT-KI showed that both produce ectodermal expression in over 96% of GFP (or RFP) positive embryos. Ectopic expression in either apical or vegetal region was observed in 5-15% of embryos (Table 1). The

rates of ectopic expression between 5P-KI and INT-KI were not statistically significantly different ($P > 0.08$, fisher's exact test, Sokal and Rohlf, 1995). Similar results were obtained with a construct that included both modules (5P-INT-KI), as expected, since the expression domains of 5P-KI and INT-KI overlap completely (Fig. 4). Though the 3P-module showed no detectable enhancer activity (Table 1, 3P-EpGFP construct), this did not preclude the possibility that it contains a repressor of ectopic expression. However, when we tested this by introduction of an INT-3P-KI construct, no statistically significant deviation in expression pattern from other constructs was observed.

5P, the module that initiates nodal expression, and its early inputs

By preventing translation of the endogenous *nodal* mRNA, the module responsible for initiation of *nodal* expression can be distinguished, since expression constructs containing this module should manifest only the early residual expression seen in the endogenous gene under conditions of MASO treatment (Fig. 2). This expression should be independent of the positive feedback input. Figure 5A shows that at 7 h the number of GFP transcripts generated by the 5P-KI construct is the same in MASO treated and in control embryos. In contrast, as seen in Fig. 5B, early expression of the INT-KI construct was essentially eliminated by MASO treatment. This pattern was consistently observed in multiple experiments (see figure legend). These data show that it is the 5P-module which is responsible for receiving the initial input.

To determine the identity of the *cis*-regulatory target sites required for the initial input, we made use of the insight from the work of Coffman et al. (2001;2004) that oral ectoderm specification is mediated by redox polarization (see Introduction). Transcription factors of the bZIP class are known to be redox sensitive (Abate et al., 1990;Amoutzias et al., 2006), and a majority of the 13 bZIP factors encoded in the sea urchin genome are expressed in early embryogenesis (Howard-Ashby et al., 2006). Therefore we examined the 5P sequence for putative bZIP binding sites, using a position weight matrix algorithm (MatInspector, see Cartharius et al., 2005). At least six sites that matched the profile of bZIP sites above default criteria were found. Because bZIP factors are known to form either homodimers or heterodimers with other bZIP factors, we focused on palindromic bZIP sites. As indicated in Fig. 6A, 5P contains two pairs of putative bZIP sites. A mutant 5P-KI-GFP construct (5P-KI-GFP xbZIP) which lacks three of the four bZIP sites was generated, and its activity was compared to that of the wild-type 5P-KI-RFP construct. To provide an internal standard of transgene expression, these constructs were co-injected, and the amount of RFP and GFP mRNA produced at each time point was measured by QPCR. As a control, a wild-type 5P-KI-RFP construct was first introduced together with the wild-type 5P-KI-GFP construct. Figure 6B shows that through 10 h the ratio of GFP to RFP is one, and thus there is no significant difference in GFP vs. RFP mRNA turnover rates that could affect the outcome of the experiment. In Fig. 6C we see the clear effect of the bZIP mutations: these almost eliminate the early expression of the 5P construct (see inset). Figure 6D shows that when in addition the positive feedback input is eliminated by MASO treatment, the difference in transcriptional activities between 5P-RFP-KI and 5P-GFP-KI_xbZIP is accentuated. These internally controlled experiments consistently produced that same pattern in two or three independent experiments (depending on time points) with very little variation (see figure legend). We conclude that the bZIP sites are indeed those through which are controlled the initial activation mediated by the 5P module. Since this is the only module recovered that displays initial activity (Fig. 5), these are most likely the sites that account for the redox-sensitive activation of the endogenous *nodal* gene.

Cis-regulatory basis of positive feedback through Nodal signaling

The experiments in Fig. 5 show that both the 5P and the INT modules are regulated by positive feedback that depends on Nodal signal generation. Nodal signals are transduced by SMAD2

and SMAD3, or orthologous factors, including in the regulation of *nodal* itself in vertebrates (Schier, 2003). The sea urchin genome encodes a single ortholog of SMAD2/3, which is expressed both maternally and zygotically during embryogenesis (Howard-Ashby et al., 2006), and this factor is the signal transducer in Nodal signaling. We identified two SMAD sites in the 5P-module and two more in the INT-module, located as indicated in Fig. 7A. A GFP construct including both modules but lacking all four of the putative SMAD sites (5P-INT-KI-GFP_xSMAD) was generated, and its transcriptional activity, as measured by QPCR, was compared to that of the wild-type construct (5P-INT-KI-GFP). Figure 7B shows that the construct bearing mutated SMAD sites fails to generate the rising expression profile of the wild-type version (or the endogenous gene) over the first few hours of expression. The same result was obtained in two or three independent experiments depending on time points (see figure legend). Since the *nodal* MASO experiments of Figs. 2 and 5 demonstrate that this rise is due to positive feedback on the *nodal* transcriptional regulatory system, the experiment of Fig. 7 shows that SMAD sites are indeed required for the feedback input.

In one important way the result of mutating the SMAD sites and the result of *nodal* MASO treatment differ. As we showed in Fig. 5 nodal MASO does not affect the initial activity at 7 h mediated by the 5P module, but unexpectedly, we observed that mutation of the four SMAD sites abolished 7 h expression, whether or not the bZIP sites are also mutated (see inset in Fig. 7B). The same patterns were observed in internally controlled experiments using coinjected wild-type constructs driving GFP, and mutant constructs driving RFP (data not shown). It follows that that at least one of the putative SMAD sites we tested is also required for the initial activity measured at 7 h.

Spatial repression input into the INT module

In the course of the FamilyRelations analysis, we noticed that the 3′ portion of the INT module (P6 - P7 in Fig. 3A) was detected at a 70% identity cutoff, but was lost when 80% identity was required within the 50 bp window. To examine role of this relatively less conserved region of the INT module, constructs lacking it were generated (short INT-KI; see Fig. 3D). This construct produced a dramatic increase in vegetal expression, 29.7%, vs. 6.8% for the control INT-KI construct (Table 1, $P < 0.003$). The same results were obtained if the construct included the 5P module (5P-Short INT-KI; Table1). Short INT-EpGFP gave even more extreme vegetal expression (Table 1). Nor was vegetal expression eliminated by inclusion of 5P (5P-Short INT-KI), or of a large amount of additional 5′ sequence upstream of 5P together with the 3P conserved patch as well (long 5P-short INT-3P-KI; see Fig. 3D and Table1). These results show that the slightly less conserved 3′ portion of the INT module contains target sites for a repressor of vegetal expression, and that neither the 5P-module nor the 3P-module has an effect on the vegetal repression function of the INT-module.

Discussion

Functional *cis*-regulatory dissection of the *nodal* gene reveals two conserved modules, 5P and INT, that appear to be responsible for the initiation and maintenance of *nodal* expression in the presumptive oral ectoderm. While we cannot speak to the possibility that there exist other redundant or partially redundant modules, 5P and INT are individually sufficient to explain the major features of *nodal* regulation.

Cis-regulatory logic of the nodal gene

A summary of the *cis*- and *trans*- inputs into the *nodal* regulatory system that are implied or demonstrated in this work is presented in Fig. 8.

Oral polarization depends initially on a cleavage-stage redox gradient, as established by Coffman et al. (2004). Since *nodal* expression is the initial transcriptional response required for activation of the gene regulatory network for oral ectoderm specification, initial activation of the *nodal* gene should be mediated by redox sensitive transcription factors. We show here that mutation of putative bZIP sites blocks early activation of the 5P *cis*-regulatory construct, the module capable of initiating expression in the absence of prior Nodal signaling. Thus we predict that the cytoplasmic transducers of the polarized redox state will be one of the bZIP factors known to be present in the early embryo, and that these provide the initial input into the *nodal* transcriptional system on the oral side of the embryo (step 1 in Fig. 8). bZIP factors are known to be redox sensitive (Abate et al., 1990;Amoutzias et al., 2006). A further link to them is provided by the study of Bradham and McClay (2006), who showed that p38 MAP kinase, which is an evolutionarily conserved intermediary in intracellular redox signaling, is an upstream regulator of *nodal* expression (in *L. variegatus*); the p38 kinase pathway regulates the activities of bZIP transcription factors (e.g., Inoue et al., 2005). However, there is an additional component to initial activation, as revealed by the requirement for the presence of at least one SMAD target site for initial activation, besides the bZIP sites. We do not know the mechanism for this, but since both inputs are absolutely required (Fig. 7B) the initiation system works by AND logic (Istrail and Davidson, 2005). Therefore, the second input need not be localized to the oral side. One possibility is that it derives from the underlying vegetal cells since there is some evidence that vegetal signaling is needed for oral ectodermal specification (see Introduction); another is that it derives from one of the three TGFβ factors present in the embryo at this time (unpublished data, JN and EHD), since these would also utilize SMAD site(s).

Once *nodal* expression is initiated, the Nodal signaling pathway strongly boosts *nodal* gene expression, and this feedback stimulation accounts for most of the output of the gene after very early times (input 2 in Fig. 8). Our *cis*-regulatory analysis demonstrates that the SMAD2/3 sites are indeed responsible for the positive feedback input, as would be predicted from the results of Yaguchi et al. (2007). Their study also showed that secreted Nodal diffuses to the immediately adjacent cells. Thus the positive feedback regulation within the oral ectoderm territory is likely an intercellular one, in which Nodal secreted by each given cell activates *nodal* transcription (and Nodal synthesis) in adjacent cells, though we do not wish to exclude the possibility of autocrine signaling as well.

There are three territories in which *nodal* expression is repressed, viz. the apical and vegetal territories, and the aboral ectoderm. Since Nodal does diffuse from cell to cell, and since reception of the signal can activate the INT maintenance module even without additional inputs, an active repression mechanism must protect these territories from nodal transcription. We found direct *cis*-regulatory evidence for a vegetal transcriptional repressor that operates through target sites in the distal region of the INT module (input 3 of Fig. 8). When this region is missing expression spreads to the vegetal domain, implying the existence of a yet unidentified activator functional in vegetal cells, and no doubt elsewhere, to which INT can respond. However, we have not been able to obtain reproducible evidence for an equivalent apical domain repressor. The most likely explanation is that the missing repressor sites are located in DNA sequence that fell below the threshold level of conservation in our analysis, since the vegetal repressor sites almost missed that criterion as well. A different kind of explanation may pertain to the aboral ectoderm. A plausible argument is that aboral extension of *nodal* expression is blocked by an extracellular antagonist of Nodal, Antivin/Lefty, which interferes with the interaction of Nodal with its receptors (Duboc et al., 2004;Sakuma et al., 2002;Schier, 2003). Over-expression of Lefty in fact decreases nodal expression in sea urchin embryos (Duboc et al., 2004). The gene encoding Lefty is expressed in the oral ectoderm, and this protein may diffuse peripherally to the *nodal* expression domain, thereby preventing Nodal signaling from spreading to the aboral ectoderm.

An oral ectoderm community effect

This is the second example in the early sea urchin embryo of an intra-territorial signaling function, the mechanism of which is transcriptional feedback from the signal transduction apparatus to the gene encoding the intra-territorial signaling ligand. A perfectly analogous system operates in the endomesoderm of the early embryo, where the *wnt8* gene is active intra-territorially, and transcription of this gene depends on an input from β-catenin/TCF, the product of the Wnt signal transduction system. This was shown to be a direct input into the *wnt8* gene in the *cis*-regulatory study of Minokawa et al. (2005), just as is the input into *nodal* via the SMAD sites demonstrated in the present study. In both the Wnt8 and Nodal systems a short-range signaling ligand is expressed throughout the territory, so that all cells both receive this ligand and transcriptionally express the gene encoding it. A gene regulatory network analysis of mesoderm specification in *Xenopus* (Koide et al., 2005) reveals similar mechanisms, in this case involving genes encoding two other intra-territorial ligands, viz. a BMP and an EGF ligand (reviewed by Davidson, 2006). We have named this type of regulatory subcircuit the "community effect," borrowing a term coined by Gurdon (Gurdon, 1988) for a situation in which intra-territorial signaling is required for a tissue to retain its territorial state of specification. Gene regulatory network analysis shows that in each of the cited sea urchin and *Xenopus* cases expression of additional key regulatory genes required to set up the territorial regulatory states depends directly on the same respective signal transduction inputs (for *Xenopus*,Koide et al., 2005; for sea urchin endomesoderm, Davidson, 2006; Smith et al, 2007; for sea urchin oral ectoderm, Y.-H. Su, E. H. Davidson, unpublished data). Thus community effect subcircuitry at once defines the interrelationship of cells within a transcriptional domain, and explains how maintenance of their functional linkage contributes directly to their territory-specific regulatory state.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

JN appreciates Sagar Damle, Stefan Materna, Joel Smith, Paola Oliveri, Roger Revilla-i-Domingo, Andy Ransick, and Charles Titus Brown for criticisms and suggestions; Dave McClay and Cyndi Bradham for sharing their experiences with nodal with us; and Julie Hahn, Ping Dong, and Miki Yun for technical help. This work was supported by NIH grant HD-37105.

References

Abate C, Patel L, Rauscher F J. r. Curran T. Redox regulation of fos and jun DNA-binding activity in vitro. Science 1990;249:1157–1161. [PubMed: 2118682]

Amoutzias G, Bornberg-Bauer E, Oliver S, Robertson D. Reduction/oxidation-phosphorylation control of DNA binding in the bZIP dimerization network. BMC Genomics 2006;7:107. [PubMed: 16674813]

Arnone MI, Bogarad LD, Collazo A, Kirchhamer CV, Cameron RA, Rast JP, Gregorians A, Davidson EH. Green fluorescent protein in the sea urchin: new experimental approaches to transcriptional regulatory analysis in embryos and larvae. Development 1997;124:4649–4659. [PubMed: 9409681]

Arnone, MI.; Dmochowski, IJ.; Gache, C. Using reporter genes to study cis-regulatory elements. In: Ettensohn, CA.; Wessel, GM.; Wray, GA., editors. Development of sea urchins, ascidians, and other invertebrate deuterostomes: Experimental approaches. 74. Elservier Academic Press; San Diego: 2004. p. 621-652.

Bradham CA, McClay DR. p38 MAPK is essential for secondary axis specification and patterning in sea urchin embryos. Development 2006;133:21–32. [PubMed: 16319119]

Brown CT, Xie Y, Davidson EH, Cameron RA. Paircomp, FamilyRelationsII and Cartwheel: tools for interspecific sequence comparison. BMC Bioinformatics 2005;6:70. [PubMed: 15790396]

Cameron RA, Fraser SE, Britten RJ, Davidson EH. Segregation of oral from aboral ectoderm precursors is completed at fifth cleavage in the embryogenesis of Strongylocentrotus purpuratus. Dev. Biol 1990;137:77–85. [PubMed: 2295368]

Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 2005;21:2933–2942. [PubMed: 15860560]

Coffman JA, Davidson EH. Oral-Aboral axis specification in the Sea Urchin embryo: I. Axis entrainment by respiratory asymmetry. Dev. Biol 2001;230:18–28. [PubMed: 11161559]

Coffman JA, McCarthy JJ, Dickey-Sims C, Robertson AJ. Oral-aboral axis specification in the sea urchin embryo: II. Mitochondrial distribution and redox state contribute to establishing polarity in Strongylocentrotus purpuratus. Dev. Biol 2004;273:160–171. [PubMed: 15302605]

Duboc V, Rottinger E, Besnardeau L, Lepage T. Nodal and BMP2/4 signaling organizes the oral-aboral axis of the Sea Urchin embryo. Dev. Cell 2004;6:397–410. [PubMed: 15030762]

Flowers VL, Courteau GR, Poustka AJ, Weng W, Venuti JM. Nodal/activin signaling establishes oral-aboral polarity in the early sea urchin embryo. Dev. Dyn 2004;231:727–740. [PubMed: 15517584]

Gurdon JB. A community effect in animal development. Nature 1988;336:772–774. [PubMed: 3205305]

Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH. Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus. Dev. Biol 2006;300:90–107. [PubMed: 17054934]

Inoue H, Hisamoto N, An JH, Oliveira RP, Nishida E, Blackwell TK, Matsumoto K. The C. elegans p38 MAPK pathway regulates nuclear localization of the transcription factor SKN-1 in oxidative stress response. Genes Dev 2005;19:2278–2283. [PubMed: 16166371]

Istrail S, Davidson EH. Gene Regulatory Networks Special Feature: Logic functions of the genomic cis-regulatory code. Proc. Natl. Acad. Sci. USA 2005;102:4954–4959. [PubMed: 15788531]

Koide T, Hayata T, Cho KWY. Gene Regulatory Networks Special Feature: Xenopus as a model system to study transcriptional regulatory networks. Proc. Natl. Acad. Sci. USA 2005;102:4943–4948. [PubMed: 15795378]

Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA, Copeland NG. A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 2001;73:56–65. [PubMed: 11352566]

Livant DL, Hough-Evans BR, Moore JG, Britten RJ, Davidson EH. Differential stability of expression of similarly specified endogenous and exogenous genes in the sea urchin embryo. Development 1991;113:385–398. [PubMed: 1782856]

Materna SC, Howard-Ashby M, Gray RF, Davidson EH. The C2H2 zinc finger genes of Strongylocentrotus purpuratus and their expression in embryonic development. Dev. Biol 2006;300:108–120. [PubMed: 16997293]

Minokawa T, Wikramanayake AH, Davidson EH. cis-Regulatory inputs of the wnt8 gene in the sea urchin endomesoderm network. Dev. Biol 2005;288:545–558. [PubMed: 16289024]

Revilla-i-Domingo R, Minokawa T, Davidson EH. R11: a cis-regulatory node of the sea urchin embryo gene network that controls early expression of SpDelta in micromeres. Dev. Biol 2004;274:438–451. [PubMed: 15385170]

Sakuma R, Ohnishi Y.-i. Meno C, Fujii H, Juan H, Takeuchi J, Ogura T, Li E, Miyazono K, Hamada H. Inhibition of Nodal signalling by Lefty mediated through interaction with common receptors and efficient diffusion. Genes Cells 2002;7:401–412. [PubMed: 11952836]

Schier AF. Nodal signaling in vertebrate development. Annu. Rev. Cell. Dev. Biol 2003;19:589–621. [PubMed: 14570583]

Sea Urchin Genome Sequencing Consortium. The genome of the sea urchin Strongylocentrotus purpuratus. Science 2006;314:941–952. [PubMed: 17095691]

Sokal, RR.; Rohlf, FJ. Biometry. W. H. Freeman and Company; New York: 1995.

Wikramanayake AH, Brandhorst BP, Klein WH. Autonomous and non-autonomous differentiation of ectoderm in different sea urchin species. Development 1995;121:1497–1505. [PubMed: 7789279]

Wikramanayake AH, Huang L, Klein WH. beta -Catenin is essential for patterning the maternally specified animal-vegetal axis in the sea urchin embryo. Proc. Natl. Acad. Sci. USA 1998;95:9343–9348. [PubMed: 9689082]

Wikramanayake AH, Klein WH. Multiple signaling events specify ectoderm and pattern the oral-aboral axis in the sea urchin embryo. Development 1997;124:13–20. [PubMed: 9006063]

Wong ML, Medrano JF. Real-time PCR for mRNA quantitation. BioTechniques 2005;39:75–85. [PubMed: 16060372]

Yaguchi S, Yaguchi J, Burke RD. Sp-Smad2/3 mediates patterning of neurogenic ectoderm by Nodal in the sea urchin embryo. Dev. Biol 2007;302:494–503. [PubMed: 17101124]

Yon J, Fried M. Precise gene fusion by PCR. Nucleic Acids Res 1989;17:4895. [PubMed: 2748349]

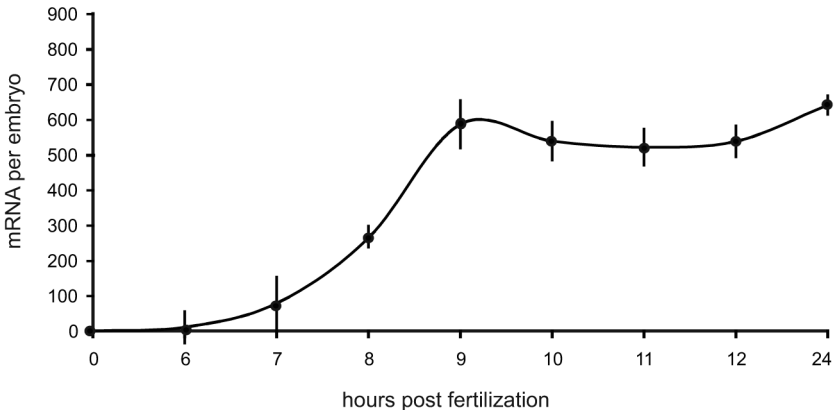

Fig 1.

Temporal expression of *nodal* measured by QPCR. The number of transcripts at each time point was averaged from three independent experiments. Vertical bars indicate normalized standard deviation (SD) (i.e., standard deviation/mean × 100). Numbers of transcripts were estimated by comparison to an internal standard, as described (Materna et al., 2006).

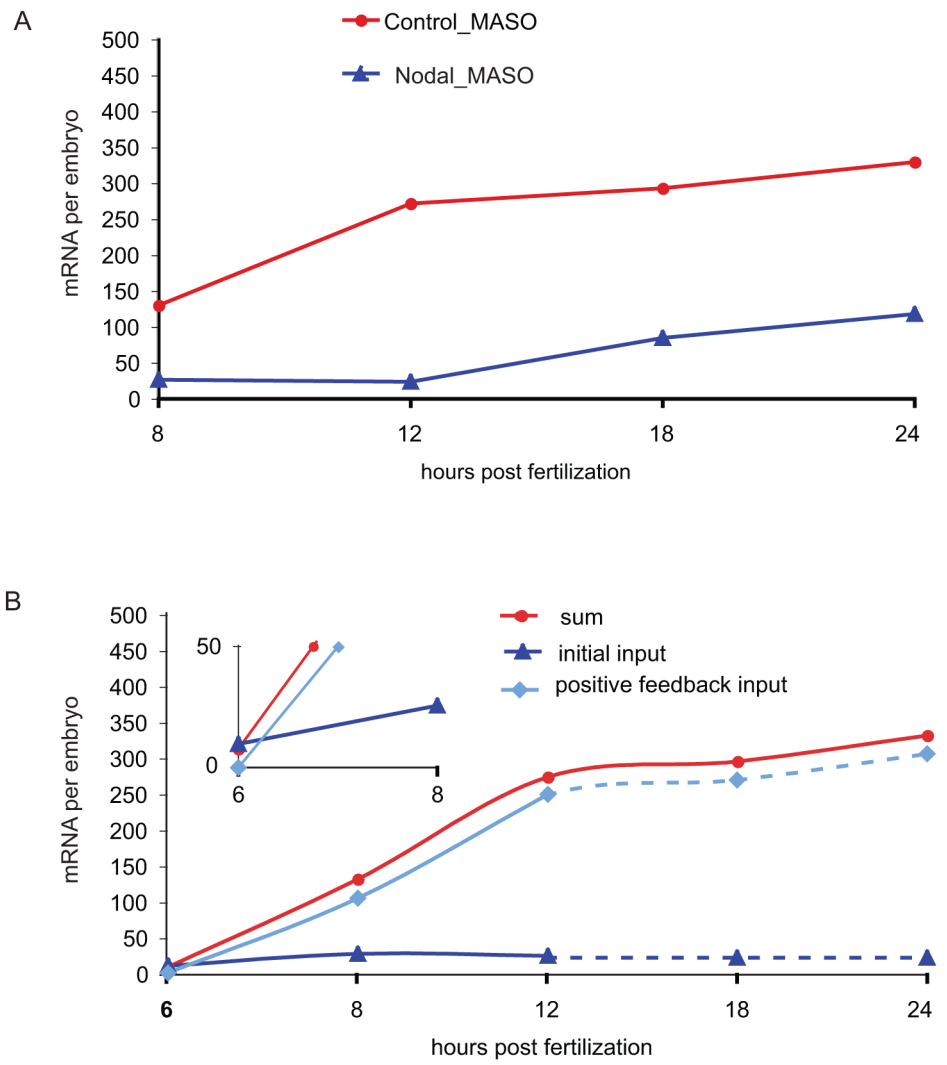

Fig 2.

Feedback control of nodal gene expression. (A) Effects of nodal MASO on endogenous output of *nodal* mRNA. Blue triangles indicate nodal transcripts in MASO injected embryos at each time point; and red circles indicate *nodal* transcripts in embryos injected with a non-specific control MASO. Measurements were averaged from two independent QPCR experiments. Between embryos respectively injected with nodal MASO and control MASO, the SD of the ratio of the numbers of *nodal* transcripts for each time point was below 15% of the mean ratio, except for 24 hpf, where SD was 48%. The SDs were estimated from five independent experiments, except for 18 hpf (n = 2). (B) Model for nodal expression combining Figs. 1 and 2A. The mRNA numbers are representative values. Expression of nodal measured in this experiment, red line, is considered as the sum of the level of expression due to the initial (and for the brief considered, continuing) input, blue line; and the level of expression due to the positive feedback input, pale blue line. The initial input must precede the positive feedback. The relative values of these components >12 h remain hypothetical, indicated by the dashed lines.

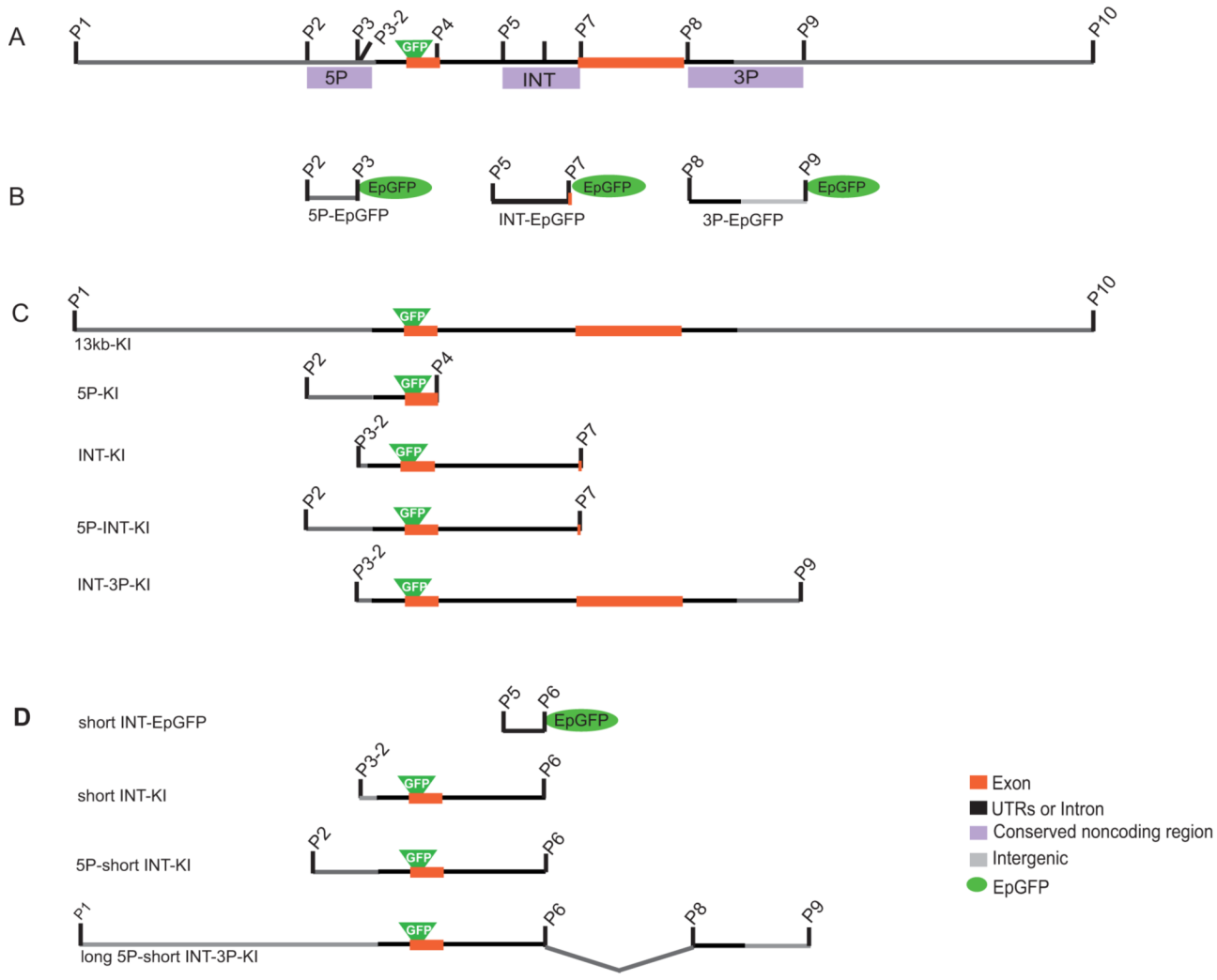

Fig 3.

Conserved noncoding regions in the *nodal* locus, and maps of expression constructs. (A) Conserved noncoding regions, and locations of primers used to generate expression constructs (P1-P10; see supplementary table S1 for the sequences of primers). The location of a GFP coding sequence inserted by recombination in exon1 is also shown (see Materials and methods). (B) Maps of reporter constructs harboring GFP cDNA and including the *endo16* basal promoter (EpGFP). (C and D) Maps of GFP knock-in (KI) constructs. Each fragment was PCR amplified with the indicated pair of flanking primers, using the GFP-BAC KI construct as template.

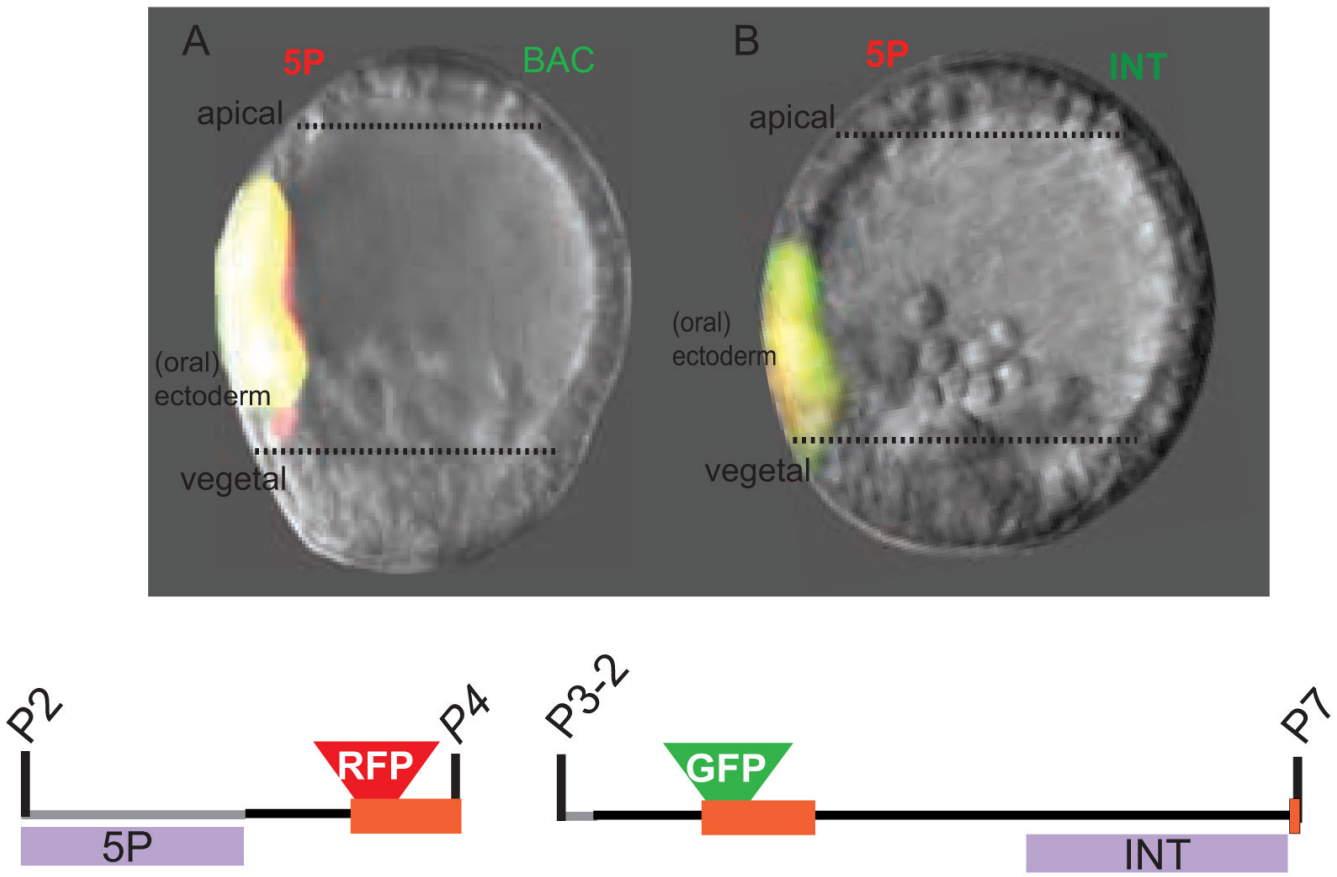

Fig 4.

Spatial activities of co-injected Nodal GFP-BAC, the 5P-module, and the INT-module in 24 h embryos. Color of expression domains produced by each construct corresponds to the color of the fluorescence generated by each reporter; yellow areas are overlapping regions of red and green. (A) Co-injected Nodal GFP-BAC (green) and 5P-KI-RFP (red) are active on the same (i.e., oral) side of the ectoderm. (B) Co-injected 5P-KI-RFP (red) and INT-KI-GFP (green) are active coincidentally. (C) Maps of 5P-KI-RFP and INT-KI-GFP constructs.

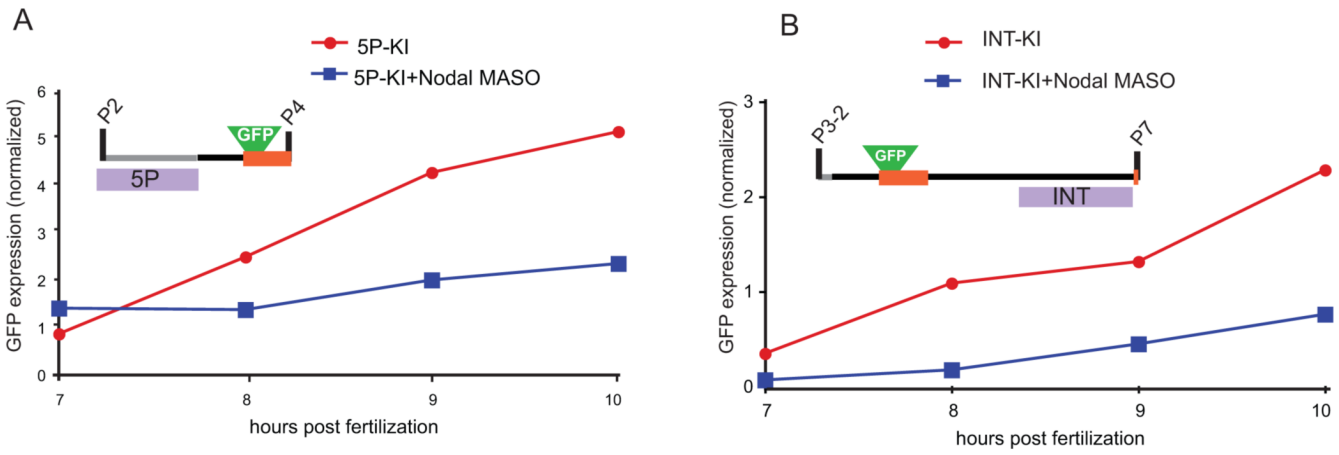

Fig 5.

Effect of *nodal* MASO on the activites of the 5P and the INT *cis*-regulatory modules. Nodal MASO was used to eliminate the positive feedback input. Quantities of GFP transcript were normalized to the amount of incorporated DNA, following the method of Revilla-i-Domingo et al. (2004). Red circles indicate numbers of GFP transcripts in control embryos and blue squares those in nodal MASO injected embryos. Results were averaged from two independent sets of experiments. (A) Effect of *nodal* MASO on the activity of the 5P-KI construct. Initial (7 h) GFP mRNA was not down regulated by *nodal* MASO, and was down regulated at later time points. The SD for the ratio of the numbers of GFP transcripts from nodal MASO injected and control embryos for each time point was below 24% of the mean ratio, except for 8 hpf (SD = 53%). The SDs were estimated from three independent experiments except for 10 hpf (n = 2). (B) Effect of *nodal* MASO on the activity of the INT-KI construct. The level of GFP transcripts driven by the INT module was depressed by *nodal* MASO at all time points considered. The SD for the ratio of the numbers of GFP transcripts from nodal MASO injected and control embryos for each time point was below 10% of the mean ratio except for 9 hpf, where SD was 45%. The SDs were estimated from two independent experiments.

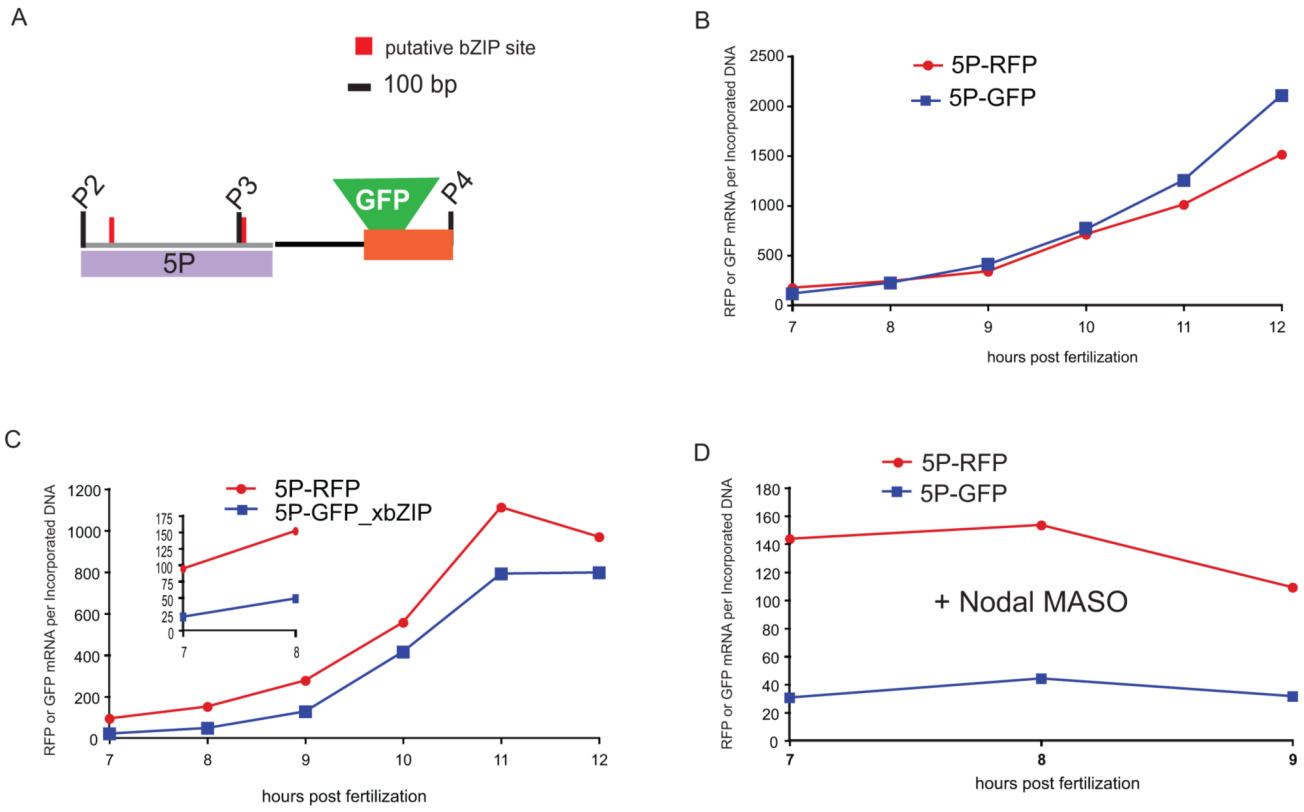

Fig 6.

Effects of bZIP site mutations on activity of the 5P module. GFP and RFP constructs were co-injected. Data shown are from one experiment; consistent results were obtained in other experiments that included fewer time points (not shown). (A) Location of two pairs of putative bZIP sites in the 5P module. (B) Control experiment comparing expression of GFP and RFP constructs driven by the same 5P *cis*-regulatory module ($SD_{GFP/RFP}$ = 23% of the mean ratio over developmental stages). (C) Comparison of the levels of GFP transcripts produced by a mutant 5P-module lacking the two pairs of bZIP site, and in the same embryos, of RFP produced by the wild-type 5P construct ($SD_{GFP/RFP}$ < 15% of the mean for 7 - 10 hpf, n = 3 for each time point). (D) Same comparison as in (C), except in the presence of nodal MASO ($SD_{GFP/RFP}$ < 10% of the mean ratio, n = 2 for each time point).

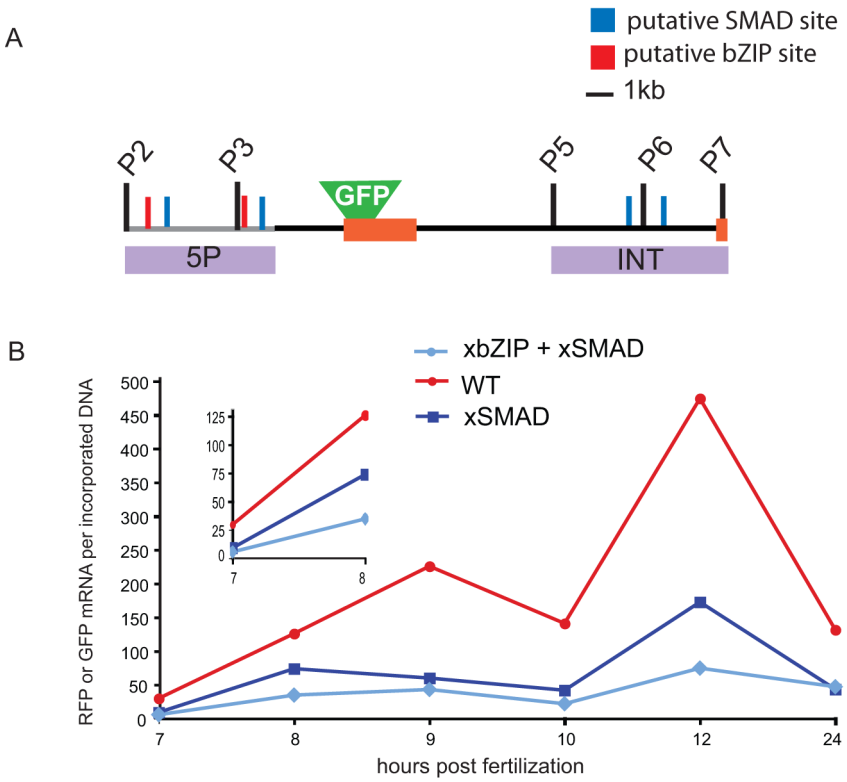

Fig 7.

Effect of mutations of putative SMAD sites in the 5P and INT modules. Data are all from the same experiment; consistent results were obtained in other experiments with fewer time points (not shown). (A) Map of 5P-INT-KI-GFP showing location of putative SMAD sites in the 5P and INT modules. (B) Transcriptional activities of the wild-type 5P-INT-KI-GFP construct, the 5P-INT-KI-GFP_xSMAD construct ($SD_{\underline{mutant}/wild\text{-}type}$ = 38% of the mean ratio for 7 hpf and $SD_{\underline{mutant}/wild\text{-}type}$ < 17% for 8 - 24 hpf, n = 3 except for 10 hpf), and the 5P-INT-KI-GFP_xSMAD_bZip construct($SD_{\underline{mutant}/wild\text{-}type}$ < 10%, n = 2 for each time point), which lacks both SMAD sites and bZIP sites.

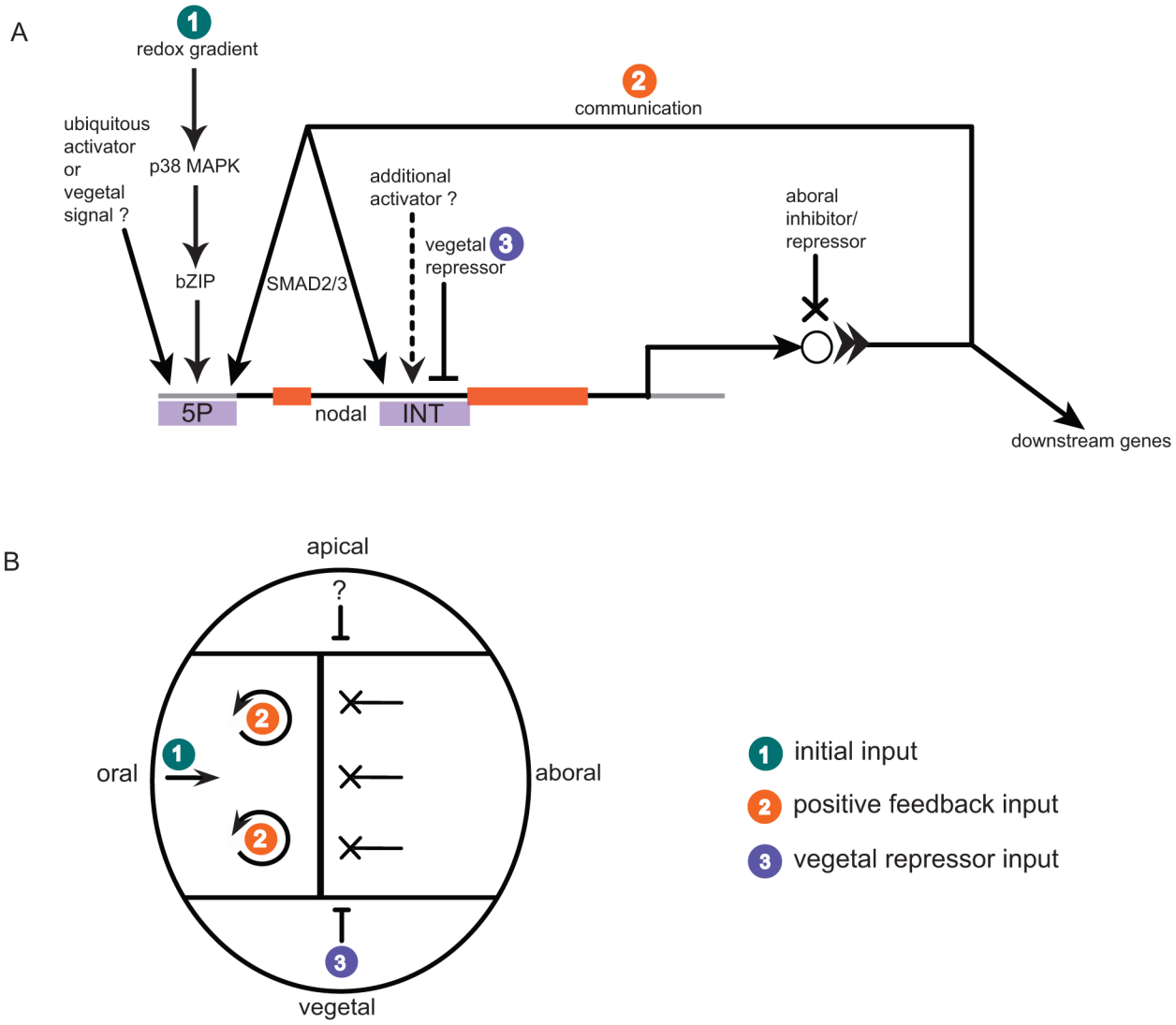

Fig 8.

A *cis*- and *trans*-regulatory model for the temporal and spatial expression of *nodal*. (A) *Cis*-regulatory inputs into the *nodal* gene and disposition of its output. The three inputs demonstrated in the present study are indicated: initial input (1), positive feedback input (2), and vegetal repressor input (3). Other implied or proposed inputs (see text) are: a ubiquitous activator or an activating input driven by a vegetal signal, mediated by a 5P SMAD site; an extracellular aboral inhibition of Nodal signaling (open circle); an apical repressor input, and the input into INT of whatever activator accounts for ectopic expression in the vegetal region in the absence of the 3′ region of INT module. (B) Locations in the embryo of the *trans* factors that convey these inputs at blastula stage.

Table 1

Spatial expression of GFP constructs

Constructs	Genomic region[a]	Ectoderm % ()[b]	Apical %	Vegetal %	Number of GFP+ embryos (%)	Number of counted embryos
5P-EpGFP	P2 - P3	91.4% (70.3%)	10.9%	18.8%	128 (55%)	231
INT-EpGFP	P5 - P7	90.9% (67.2%)	1.9%	30.9%	131 (43%)	302
3P-EpGFP	P8 - P9	-	-	-	Almost no GFP expression	> 100
5P-KI	P2 - P4	96.5% (74.6%)	16.9%	9.5%	87 (49%)	178
INT-KI	P3 - P7	98.7% (85.3%)	4.3%	10.4%	82 (31%)	265
5P-INT-KI	P2 - P7	97.7% (68.0%)	17.9%	14.1%	131 (63%)	209
INT-3P-KI[c]	P3 - P9	98.3% (79.7%)	13.5%	6.8%	59 (60%)	99
Short INT-KI[c]	P3 - P6	98.3% (57.6%)	12.7%	29.7%	59 (62%)	95
Short INT-EpGFP[c]	P5 - P6	64.3% (31%)	3.6%	65.5%	42 (43)	97
5P-short INT-KI	P2 - P6	81.6% (60.5%)	11.9%	27.6%	176 (72%)	243
Long 5P-short INT-3P-KI	P1 - P6 + P8 - P9	88.8% (62.9%)	11.9%	25.2%	143 (51%)	279

[a] See Fig. 3A for primer positions

[b] Expression exclusively in ectoderm

[c] Experiment was done only once